7-2019

# How to Gauge a Combination of Uncertainties of Different Type: General Foundations

Ingo Neumann
*Leibniz University Hannover*, neumann@gih.uni-hannover.de

Vladik Kreinovich
*University of Texas at El Paso*, vladik@utep.edu

Thach N. Nguyen
*Banking University of Ho Chi Minh City*, Thachnn@buh.edu.vn



Comments:
Technical Report: UTEP-CS-19-73

# How to Gauge a Combination of Uncertainties of Different Type: General Foundations

Ingo Neumann[1], Vladik Kreinovich[2], and Thach Ngoc Nguyen[3]

[1]Geodaetisches Institut
Leibniz Universitaet Hannover
Nienburger Strasse 1
30167 Hannover, Germany
neumann@gih.uni-hannover.de
[2]Department of Computer Science
University of Texas at El Paso
500 W. University
El Paso, TX 79968, USA
vladik@utep.edu
[3]Banking University of Ho Chi Minh City
56 Hoang Dieu 2, Quan Thu Duc, Thu Duc
Ho Chi Minh City, Vietnam, Thachnn@buh.edu.vn

**Abstract**

In many practical situations, for some components of the uncertainty (e.g., of the measurement error) we know the corresponding probability distribution, while for other components, we know only upper bound on the corresponding values. To decide which of the algorithms or techniques leads to less uncertainty, we need to be able to gauge the combined uncertainty by a single numerical value – so that we can select the algorithm for which this values is the best. There exist several techniques for gauging the combination of interval and probabilistic uncertainty. In this paper, we consider the problem of gauging the combination of different types of uncertainty from the general fundamental viewpoint. As a result, we develop a general formula for such gauging – a formula whose particular cases include the currently used techniques.

# 1 Formulation of the Problem

**Need to gauge uncertainty.** Measurements are never absolutely accurate, the measurement result $\widetilde{x}$ is, in general, different from the actual (unknown) value $x$ of the corresponding quantity. To understand how accurate is the measurement, we need to gauge the corresponding uncertainty, i.e., to provide a number describing the corresponding measurement error $\Delta x \stackrel{\text{def}}{=} \widetilde{x} - x$. For different types of uncertainty, it is natural to use different characteristics. For example:

- For probabilistic uncertainty, when we know the probability distribution of the corresponding measurement error, a natural measure of deviation is the standard deviations $\sigma$.
- For interval uncertainty, we only know the upper bound $\Delta$ on the absolute value of the measurement error; this upper bound is an appropriate measure of uncertainty.

It is reasonable to select a characteristic that is described in the same unit as the measured quantity itself. In this case, if we change the measuring unit to the one which is $\lambda$ times smaller, then:

- not only all numerical value $x$ should multiply by $\lambda$ ($x \to x' = \lambda \cdot x$), but also
- the corresponding characteristic of uncertainty should change the same way: $u \to u' = \lambda \cdot c$.

Similarly, if we simply change the sign of the quantity – which, for many quantities like coordinate or charge, does not change its physical sense – then the corresponding characteristic of uncertainty should not change: $u' = u$.

In general, when we go from $x$ to $x' = c \cdot x$, then the corresponding characteristic of uncertainty should change as $u' = |c| \cdot u$.

**Need to combine uncertainty and to gauge the combined uncertainty.** The measurement error often consists of several components:

$$\Delta x = \Delta x_1 + \ldots + \Delta x_k.$$

For each of these components $\Delta x_i$, we usually know the corresponding characteristic of uncertainty $u_i$. Based on these characteristics, we need to estimate the characteristic $u$ of the overall uncertainty $\Delta x$.

A similar problem occurs when we process data, i.e., when, based on the measurement results $\widetilde{x}_i$, we compute the value of some auxiliary quantity $y$ depending on $x$ in known way, as $y = f(x_1, \ldots, x_n)$, for some algorithm $f(x_1, \ldots, x_n)$. To estimate $y$, we use the measurement results $\widetilde{x}_i$ and thus, come up with an

estimate $\widetilde{y} = f(\widetilde{x}_1, \ldots, \widetilde{x}_n)$. We need to estimate the resulting approximation error

$$\Delta y = \widetilde{y} - y = f(\widetilde{x}_1, \ldots, \widetilde{x}_n) - f(x_1, \ldots, x_n) =$$
$$f(\widetilde{x}_1, \ldots, \widetilde{x}_n) - f(\widetilde{x}_1 - \Delta x_1, \ldots, \widetilde{x}_n - \Delta x_n).$$

Measurements are usually reasonably accurate, so the measurement errors $\Delta x_i$ are small, and thus, we can safely ignore terms which are quadratic or of higher order in terms of $\Delta x_i$ and consider only the linear terms. Then, $\Delta y = \sum_{i=1}^{k} c_i \cdot \Delta x_i$, where $c_i$ is the value of the corresponding partial $c_i \stackrel{\text{def}}{=} \dfrac{\partial f}{\partial x_i}$ computed as the point $(\widetilde{x}_1, \ldots, \widetilde{x}_n)$. Once we know the uncertainty characteristics $u_i$ of each measurement error $\Delta x_i$, we can find the uncertainty characteristics $U_i = |c_i| \cdot u_i$ of each term $X_i = c_i \cdot \Delta x_i$. Based on these characteristics, we need to estimate the uncertainty characteristic of the sum $\Delta y = X_1 + \ldots + X_n$.

The need to characterize the joint uncertainty by a single number comes from the desire to select a single less uncertain option. For example, in the traditional Markowitz's portfolio allocation problem (see, e.g., [3]), when we have full information about all the probabilities, the objective is to find, among all portfolios with the given value of expected rate of return, the one with the smallest possible standard deviation I(which, in this case, corresponds to the smallest possible risk). In many practical situations, we know probabilities only with some uncertainty. As a result, for each portfolio, in addition to the random uncertainty, we have an additional uncertainty caused by the fact that we only have partial knowledge about the corresponding probabilities.

- If we minimize the random component, we risk missing a huge interval component.
- If we minimize the interval component, we risk missing a huge random component.

It is more adequate to minimize the appropriate combination of both uncertainties, this will make sure that none of the components become too large.

**How uncertainty is combined and gauged now.** In the case of probabilistic uncertainty, if we know the standard deviations $\sigma_i$ of each component $\Delta x_i$ and we have no information about their correlation, a natural idea is to assume that the error components are independent. The same conclusion can be made if we use the Maximum Entropy approach, which recommend to select, among all possible joint distributions, the one with the largest possible value of entropy [2]. It is well known that the variance of the sum of several independent random variables is equal to the sums of their variances, so for the variance $\sigma^2$ of the sum $\Delta x$ we have $\sigma^2 = \sigma_1^2 + \ldots + \sigma_n^2$ and $\sigma = \sqrt{\sigma_1^2 + \ldots + \sigma_n^2}$.

On the other hand, if we know that for each $i$, the component $\Delta x_i$ can take any value from the interval $[-\Delta_i, \Delta_i]$, then the largest possible value $\Delta$ of the

sum is attained when each of the components $\Delta x_i$ attains its largest possible value $\Delta_i$, so we have $\Delta = \Delta_1 + \ldots + \Delta_n$.

In these two cases, we have two different formulas for combining uncertainty: if we know the uncertainty characteristics $u_i$ of the components, then the uncertainty characteristic $u$ of the sum is equal:

- in the first case, to $u = \sqrt{u_1^2 + \ldots + u_n^2}$ and
- in the second case to $u = u_1 + \ldots + u_n$.

**What is a general case?** We are looking for the binary combination operation $u * u'$ which has the following properties:

- $u * 0 = u$, meaning that adding 0 should not change anything, including the accuracy;
- the sum does not depend on the order in which we add the components, so the result of combination should also not depend on the order in which we combine the components; so, we should have $u * u' = u' * u$ (commutativity) and $u * (u' * u'') = (u * u') * u''$ (associativity);
- monotonicity: if we replace one of the components with a less accurate one (with larger $u$), the result cannot become more accurate: if $u_1 \le u_2$ and $u_1' \le u_2'$, then we have $u_1 * u_1' \le u_2 * u_2'$.

It turns out (see, e.g., [1]) that under these conditions, every combination operation has:

- either the from $u * u' = (u^p + (u')^p)^{1/p}$ for some $p > 0$,
- or the form $u * u' = \max(u, u')$ (that corresponds to the limit case $p \to \infty$).

**Remaining problem: what is we combine uncertainties of different type?** In many cases, we have different information about the uncertainty of different components. For example, the measurement error is often represented as the sum of a systematic error (the mean value) and the remaining part which is known as a random error. About the random error component, we usually know the standard deviation, so it can be a viewed as a probabilistic uncertainty; see, e.g., [6]. However, about the systematic error component, we only know the error bound – so it is the case of interval uncertainty. How should we gauge the result of combining uncertainties of different type?

**How the combination of uncertainties of different type is gauged now?** There are several ways to gauge the combination of probabilistic and interval uncertainty.

The first way takes into account that, in practice, probability distributions are often either Gaussian (normal) or close to Gaussian [4, 5]. This empirical

fact is easy to explain: in many cases, the measurement error is a result of a large number of independent small factors, and it is known that the distribution of the sum of the large number of small independent random variables is close to Gaussian (and tends to Gaussian when the number of components tends to 0) – this fact is known as the *Central Limit Theorem* (see, e.g., [7]).

Strictly speaking, a normally distributed random variable with 0 mean can take arbitrarily large value – since its probability density function $\rho(x)$ remains positive for all values $x$. However, from the practical viewpoint, the probabilities of very large values are so small that, for all practical purposes, they can be safely ignored. Thus, in practice, we assume that all the values of a normal random variable with 0 mean and standard deviation $\sigma$ are located in an interval

$$[-k_0 \cdot \sigma, k_0 \cdot \sigma],$$

where $k_0$ depends on how small the probability we can ignore; usually, people take $k_0$ equal to 2 (corresponding to 5%), 3 (0.1%) and 6 ($10^{-6}$%).

So, a random error component with standard deviation $\sigma$ implies that this component lies in the interval $[-k_0 \cdot \sigma, k_0 \cdot \sigma]$. So, all we have to do to combine it with the interval uncertainty $[-\Delta, \Delta]$ is to combine the two intervals, and get

$$\Delta + k_0 \cdot \sigma.$$

Another frequently used approach is based on the Maximum Entropy idea, according to which, if we do not know the exact distribution, then, out of all possible probability distributions, we should select the one whose entropy is the largest; see, e.g., [2]. For example, if all we know that the systematic error is located on the interval $[-\Delta, \Delta]$, then, out of all possible probability distributions on this interval, we should select the one whose entropy is the largest – which turns out to be the uniform distribution on this interval. One can easily find that for the uniform distribution on the interval $[-\Delta, \Delta]$, the standard deviation is equal to $\frac{\Delta}{\sqrt{3}}$. Thus, to combine it with the random error component with known standard deviation $\sigma$, it is sufficient to use the general formula for combining standard deviations, and get $\sqrt{\frac{\Delta^2}{2} + \sigma^2}$.

**Need for a general approach – and what we do in this paper.** So what is the general formula? This is a problem to which, in this paper, we provide an answer.

## 2 Definitions and the Main Result

Let us assume that $\mathcal{T}$ is the set of possible types of uncertainty with $T$ elements. For simplicity, let us enumerate the types, i.e., let us identify $\mathcal{T}$ with the set $\{1, 2, \ldots, T\}$.

By combining uncertainties from some subset $S \subseteq \mathcal{T}$, we get, in effect, a new type of uncertainty. Thus, we has, in effect, as many types of uncertainty as there are nonempty subsets $S \subseteq \mathcal{T}$. Since uncertainties can be of different type, in order to properly combine them, we need to know the type. Thus, an uncertainty is described not just by a number, but also by a type.

**Definition 1.** *Let a finite set $\mathcal{T}$ be given. By an* uncertainty*, we mean a pair $(u, S)$, where:*

- *$u$ is a non-negative real number and*
- *$S$ is a non-empty subset of the set $\mathcal{T}$.*

**Definition 2.** *Let a finite set $\mathcal{T}$ be given. By a* combination operation*, we means a binary operation $*$ that maps two uncertainties $(u, S)$ and $(u', S')$ into a new uncertainty $(u'', S \cup S')$ and that has the following properties:*

- *the operation $*$ is commutative and associative;*
- *the operation $*$ is monotonic in the following sense: if $u_1 \le u_2$ and $u_1' \le u_2'$, then we have $u_1'' \le u_2''$, where $(u_1, S) * (u_1', S') = (u_1'', S \cup S')$ and*

$$(u_2, S) * (u_2', S') = (u_2'', S \cup S');$$

- *scale-invariance: for every $\lambda > 0$, if $(u, S) * (u', S') = (u'', S \cup S')$, then*

$$(\lambda \cdot u, S) * (\lambda \cdot u', S') = (\lambda \cdot u'', S \cup S');$$

- *zero-property: for each set $S$, we have $(u, S) * (0, S) = (u, S)$; and*
- *non-zero property: if $u > 0$ and $(u, S) * (u', S') = (u'', S \cup S')$, then $u'' > 0$.*

**Proposition 1.** *For each combination operation, there exist positive values $c_1, \ldots, c_T$ such that:*

- *either $(u_1, \{1\}) * (u_2, \{2\}) * \ldots * (u_T, \{T\}) = ((c_1^p \cdot u_1^p + \ldots + c_T^p \cdot u_T^p)^{1/p}, \mathcal{T})$ for all $u_i$*
- *or $(u_1, \{1\}) * (u_2, \{2\}) * \ldots * (u_T, \{T\}) = (\max(c_1 \cdot u_1, \ldots, c_T \cdot u_T), \mathcal{T})$ for all $u_i$.*

**Proof.** Due to the zero property, we have

$$(u, \mathcal{T}) \stackrel{\text{def}}{=} (u_1, \{1\}) * (u_2, \{2\}) * \ldots * (u_T, \{T\}) =$$

$$(u_1, \{1\}) * (0, \{1\}) * \ldots * (0, \{1\}) * \ldots * (u_T, \{T\}) * (0, \{T\}) * \ldots * (0, \{T\}),$$

where each term $(0, \{t\})$ is repeated $T$ times. Due to associativity and commutativity, we can rearrange the terms and get

$$(u, \mathcal{T}) = (u'_1(u_1), \mathcal{T}) * \ldots * (u'_T(u_T), \mathcal{T}),$$

where we denoted

$$(u'_t(u_t), \mathcal{T}) \stackrel{\text{def}}{=} (u_t, \{t\}) * (0, \{1\}) * \ldots * (0, \{T\}).$$

Due to non-zero property, if $u_t = 1$, then $u'_t(1) \neq 0$. Let us denote $c_t \stackrel{\text{def}}{=} u'_t(1) > 0$. Then, due to scale-invariance, we have $u'_t(u_t) = c_t \cdot u_t$ and thus,

$$(u, \mathcal{T}) = (c_1 \cdot u_1, \mathcal{T}) * \ldots * (c_T \cdot u_T, \mathcal{T}).$$

For the values of type $\mathcal{T}$, we get the usual properties of the combination operation from [1], so we conclude that for uncertainties of this type, we have either $(u, \mathcal{T}) * (u', \mathcal{T}) = ((u^p + (u')^p)^{1/p}, \mathcal{T})$ or $(u, \mathcal{T}) * (u', \mathcal{T}) = (\max(u, u'), \mathcal{T})$. In both cases, we get exactly the formulas from the proposition, The proposition is thus proven.

*Comment.* As expected, both existing methods for combining uncertainty are particular cases of this general approach – corresponding to $p = 1$ and $p = 2$.

## Acknowledgments

This work was supported by the Institute of Geodesy, Leibniz University of Hannover. It was also supported in part by the US National Science Foundation grants 1623190 (A Model of Change for Preparing a New Generation for Professional Practice in Computer Science) and HRD-1242122 (Cyber-ShARE Center of Excellence). This paper was written when V. Kreinovich was visiting Leibniz University of Hannover.

## References

[1] K. Autchariyapanitkul, O. Kosheleva, V. Kreinovich, and S. Sriboonchitta, "Quantum econometrics: how to explain its quantitative successes and how the resulting formulas are related to scale invariance, entropy, and fuzziness", In: V.-N. Huynh, M. Inuiguchi, D.-H. Tran, and Th. Denoeux (eds.), *Proceedings of the International Symposium on Integrated Uncertainty in Knowledge Modelling and Decision Making IUKM'2018*, Hanoi, Vietnam, March 13–15, 2018.

[2] E. T. Jaynes and G. L. Bretthorst, *Probability Theory: The Logic of Science*, Cambridge University Press, Cambridge, UK, 2003.

[3] H. M. Markowitz, "Portfolio selection", *The Journal of Finance*, 1952, Vol. 7, No. 1, pp. 77–91.

[4] P. V. Novitskii and I. A. Zograph, *Estimating the Measurement Errors*, Energoatomizdat, Leningrad, 1991 (in Russian).

[5] A. I. Orlov, "How often are the observations normal?", *Industrial Laboratory*, 1991, Vol. 57, No. 7, pp. 770–772.

[6] S. G. Rabinovich, *Measurement Errors and Uncertainties: Theory and Practice*, Springer, New York, 2005.

[7] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman and Hall/CRC, Boca Raton, Florida, 2011.